

a transition from B-DNA to A-DNA (27), and an en masse transition of DNA from B-form to A-form can be induced in bacterial cells, suggesting that the A-form plays an unrecognized role in stabilizing DNA under adverse conditions such as desiccation (28). Sequence analysis and structural comparison (with PDB ID 2Z3X) do not show obvious homology, suggesting that these similarities could have arisen as a result of convergent evolution.

REFERENCES AND NOTES

1. D. Prangishvili, *Annu. Rev. Microbiol.* **67**, 565–585 (2013).
2. D. Prangishvili, R. A. Garrett, E. V. Koonin, *Virus Res.* **117**, 52–67 (2006).
3. M. Krupovic, D. Prangishvili, R. W. Hendrix, D. H. Bamford, *Microbiol. Mol. Biol. Rev.* **75**, 610–635 (2011).
4. M. Krupovic, M. F. White, P. Forterre, D. Prangishvili, *Adv. Virus Res.* **82**, 33–62 (2012).
5. D. Prangishvili, T. E. Quax, *Curr. Opin. Microbiol.* **14**, 315–320 (2011).
6. D. Prangishvili et al., *Genetics* **152**, 1387–1396 (1999).
7. C. Jaubert et al., *Open Biol.* **3**, 130010 (2013).
8. M. Lübben, G. Schäfer, *J. Bacteriol.* **171**, 6106–6116 (1989).
9. C. Baker-Austin, M. Dopson, *Trends Microbiol.* **15**, 165–171 (2007).
10. A. Bize et al., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11306–11311 (2009).
11. E. R. Quemin et al., *J. Virol.* **87**, 13379–13385 (2013).
12. E. H. Egelman, *Ultramicroscopy* **85**, 225–234 (2000).
13. T. Y. Yu, J. Schaefer, *J. Mol. Biol.* **382**, 1031–1042 (2008).
14. S. A. Overman et al., *Biospectroscopy* **4** (suppl.), S47–S56 (1998).
15. W. Earnshaw, S. Casjens, S. C. Harrison, *J. Mol. Biol.* **104**, 387–410 (1976).
16. F. S. DiMaio et al., *Nat. Methods* **12**, 361–365 (2015).
17. B. R. Szymczyna et al., *Structure* **17**, 499–507 (2009).
18. Y. Song et al., *Structure* **21**, 1735–1742 (2013).
19. X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **31**, 5108–5121 (2003).
20. P. Setlow, *Trends Microbiol.* **15**, 172–180 (2007).
21. B. Setlow, P. Setlow, *J. Bacteriol.* **177**, 4149–4151 (1995).
22. C. S. Hayes, P. Setlow, *J. Bacteriol.* **179**, 6020–6027 (1997).
23. C. S. Hayes, B. Illades-Aguilar, L. Casillas-Martinez, P. Setlow, *J. Bacteriol.* **180**, 2694–2700 (1998).
24. C. S. Hayes, E. Alarcon-Hernandez, P. Setlow, *J. Biol. Chem.* **276**, 2267–2275 (2001).
25. C. S. Hayes, Z. Y. Peng, P. Setlow, *J. Biol. Chem.* **275**, 35040–35050 (2000).
26. K. S. Lee, D. Bumbaca, J. Kosman, P. Setlow, M. J. Jedrzejewski, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2806–2811 (2008).
27. S. C. Mohr, N. V. Sokolow, C. M. He, P. Setlow, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 77–81 (1991).
28. D. R. Whelan et al., *J. R. Soc. Interface* **11**, 20140454 (2014).
29. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM035269 (to E.H.E.). The map and model have been deposited in the Electron Microscopy Data Bank (accession number EMD-6310) and in PDB (ID 3J9X), respectively. We thank K. Dryden for assistance with the microscopy.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6237/914/suppl/DC1
Materials and Methods
Figs. S1 to S5
References (30–37)

3 December 2014; accepted 21 April 2015
10.1126/science.aaa4181

RNA STRUCTURE

Structure of the HIV-1 RNA packaging signal

Sarah C. Keane,¹ Xiao Heng,^{1*} Kun Lu,^{1†} Siarhei Kharytonchyk,² Venkateswaran Ramakrishnan,¹ Gregory Carter,¹ Shawn Barton,¹ Azra Husic,¹ Alyssa Florwick,¹ Justin Santos,¹ Nicholas C. Bolden,¹ Sayo McCowin,¹ David A. Case,³ Bruce A. Johnson,⁴ Marco Salemi,⁵ Alice Telesnitsky,^{2‡} Michael F. Summers^{1‡}

The 5' leader of the HIV-1 genome contains conserved elements that direct selective packaging of the unspliced, dimeric viral RNA into assembling particles. By using a ²H-edited nuclear magnetic resonance (NMR) approach, we determined the structure of a 155-nucleotide region of the leader that is independently capable of directing packaging (core encapsidation signal; Ψ^{CES}). The RNA adopts an unexpected tandem three-way junction structure, in which residues of the major splice donor and translation initiation sites are sequestered by long-range base pairing and guanines essential for both packaging and high-affinity binding to the cognate Gag protein are exposed in helical junctions. The structure reveals how translation is attenuated, Gag binding promoted, and unspliced dimeric genomes selected, by the RNA conformer that directs packaging.

Assembly of HIV-1 particles is initiated by the cytoplasmic trafficking of two copies of the viral genome and a small number of viral Gag proteins to assembly sites on the plasma membrane (1–6). Unspliced, dimeric genomes are efficiently selected for packaging from a cellular milieu that includes a substantial excess of nonviral messenger RNAs (mRNAs) and more than 40 spliced viral mRNAs (7, 8). RNA signals that direct packaging are located primarily within the 5' leader of the genome and are recognized by the nucleocapsid (NC) domains of Gag (4). Transcriptional activation, splicing, and translation initiation are also dependent on elements within the 5' leader, the most conserved region of the genome (9), and there is evidence that these and other activities are temporally

modulated by dimerization-dependent exposure of functional signals (6, 10–13).

Understanding the RNA structures and mechanisms that regulate HIV-1 5' leader function has its basis in phylogenetic, biochemical, nucleotide reactivity, and mutagenesis studies (4). The dimeric leader selected for packaging appears to adopt a highly branched secondary structure, in which there are structurally discrete hairpins and helices that promote transcriptional activation (TAR), transfer RNA (tRNA) primer binding (PBS), packaging (ψ), dimer initiation (DIS), splicing (ψ), and dimer stability (U5:AUG) (4, 14) (Fig. 1). Although nuclear magnetic resonance (NMR) signals diagnostic of TAR, PBS, ψ , DIS, U5:AUG, and polyadenylate [poly(A)] helices have been observed in spectra obtained for the full-length dimeric

leader (13, 15) (Fig. 1A), signals diagnostic of a putative SD hairpin have not been detected (colored magenta in Fig. 1A) (15), and there is little agreement among more than 20 different structure predictions for residues adjacent to the helices (4). For example, predictions vary for stretches of residues shown by in vivo nucleotide reactivity (16) and cross-linking with immunoprecipitation (17) to reside at or near sites of Gag binding (4). The TAR, poly(A), and PBS hairpins of the HIV-1 leader are not required for efficient encapsidation (15), and a minimal HIV-1 packaging element, the core encapsidation signal (Ψ^{CES}), exhibits NC binding properties and NMR spectral features similar to those of the intact 5' leader and is independently capable of directing vector RNAs into viruslike particles (15). To gain insights into the mechanism of HIV-1 genome selection, we determined the structure of Ψ^{CES} by NMR.

Contributions of slow molecular rotational motion to NMR relaxation were minimized by substituting the dimer-promoting GC-rich loop of the Ψ^{CES} DIS hairpin by a GAGA tetraloop (Fig. 1A). This prevented dimerization (Fig. 1B) but did not affect NC binding (Fig. 1C) or nuclear Overhauser effect spectroscopy (NOESY) NMR spectral patterns (18), indicating that the modified RNA retains

¹Howard Hughes Medical Institute (HHMI) and Department of Chemistry and Biochemistry, University of Maryland Baltimore County (UMBC), 1000 Hilltop Circle, Baltimore, MD 21250, USA. ²Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-5620, USA. ³Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA. ⁴One Moon Scientific, Incorporated, 839 Grant Avenue, Westfield, NJ 07090, USA, and City University of New York (CUNY) Advanced Science Research Center, 85 St. Nicholas Terrace, New York, NY 10031, USA. ⁵Department of Pathology, Immunology, and Laboratory Medicine, College of Medicine, and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA.

*Present address: Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA. †Present address: Regeneron Pharmaceuticals, Incorporated, 777 Old Saw Mill River Road, Tarrytown, NY 10591, USA. ‡Corresponding author. E-mail: summers@hhmi.umbc.edu (M.F.S.); ateles@umich.edu (A.T.)

the structure of the native dimer. Nonexchangeable aromatic and ribose H_1' , H_2' , and H_3' 1H NMR signals were assigned for nucleotides of the U5:AUG, lower-PBS, DIS, and ψ helices by sequential residue analysis of two-dimensional (2D) NOESY spectra obtained for nucleotide-specific 2H -labeled samples (18–20) (Fig. 1D). Very-long-range A-H₂ NOEs (1H - 1H distances up to ~ 7 Å) were detected in spectra of highly deuterated samples (Fig. 1E) [as observed for proteins (21)], facilitating assignments.

NMR signals that could not be assigned by nucleotide-specific labeling were identified by a fragmentation-based segmental 2H -labeling approach that we developed, in which differentially labeled 5' and 3' fragments of Ψ^{CES} were prepared separately and noncovalently annealed (Fig. 2, A and B, and fig. S1). The dimer-promoting loop of the DIS hairpin served as the fragmentation site and was substituted by a short stretch of intermolecular G:C base pairs (Fig. 2A). Differential 2H labeling afforded the following fragment-annealed RNAs (fr- Ψ^{CES} ; denoted 5' fragment:3' fragment- Ψ^{CES} ; D, perdeuterated fragment; superscripts denote sites of protonation, all other sites deuterated; e.g., G, fully protonated guanines, A^{2r}, adenosines protonated at C₂ and ribose carbons): A^{2r}:U^r- Ψ^{CES} , A^{2r}C^r:U^r- Ψ^{CES} , C^r:A^{2r}C^r- Ψ^{CES} , D:A^{2r}C^r- Ψ^{CES} , A:D- Ψ^{CES} , and D:A- Ψ^{CES} (fig. S1). Except for residues at the sites of substitution, the NMR spectra of the fr- Ψ^{CES} RNAs were consistent with those of the parent, nonfragmented RNA. For example, NOEs that correlate A124-H₂ with cytosine and uridine H_1' protons in 2D NOESY spectra obtained for nonfragmented A^{2r}C^r- Ψ^{CES} , A^{2r}U^r- Ψ^{CES} , and A^{2r}C^rU^r- Ψ^{CES} samples were also detected in spectra obtained for fragment-annealed A^{2r}:U^r- Ψ^{CES} and A^{2r}C^r:U^r- Ψ^{CES} constructs, indicating that A124 resides near a cytosine (C125) in the 5' fragment and a uridine (U295) in the 3' fragment (Fig. 2C). More than 80 long-range and sequential A-H₂ NOEs were identified by using the 2H -edited NMR approach (Fig. 2E). The 1H NMR assignments were validated by NOE cross peak pattern redundancy and database chemical shift analyses (18, 22) (fig. S2).

The NMR data indicate that residues proximal to the major splice donor site do not form a hairpin but instead participate in long-range base pairing within an extended DIS stem and a short helical segment, H1 (Fig. 2E). To determine whether this secondary structure is also adopted by the native 5' leader, we obtained NOESY data for dimeric, 2H -labeled 5' leader constructs. Adenosine-H₂ signals diagnostic of the U5:AUG, DIS, PBS, and Ψ helices were observed in spectra obtained for the native leader ([5'-L₂]), as expected (15). However, signals diagnostic of H1 were only detectable upon removal of the upper PBS loop (substituted by a GAGA tetraloop; [5'-L^{APBS}]), which eliminated broad upper PBS signals that overlapped with the A124-H₂ signal of H1 (Fig. 2D). This construct exhibits dimerization, NC binding, and NMR properties similar to those of the intact leader (15) and directs both noncompetitive (15) and competitive RNA packaging with near-wild-type efficiency [94 ± 4% and 93 ± 18%, respectively

(reported as mean ± standard deviation)] (Fig. 2F). Thus, the secondary structure observed for Ψ^{CES} , including the H1 helix, is also adopted by the 5' leader.

NOE-restrained structure calculations (18) reveal that Ψ^{CES} adopts a tandem three-way junction structure (Fig. 3, A to C, and fig. S3). The overall shape is quasi-tetrahedral, with the U5:AUG, H1, and ψ helices forming a plane that is nearly perpendicular to the plane formed by the H1, PBS, and DIS helices (Fig. 3A). Splice-site residues G289 and G290 are base-paired with C229 and U228, respectively; adjacent residues are base-paired within or near the H1-PBS-DIS (three-way-2) junction (Fig. 3, B to D); and residues of

AUG are base-paired within the U5:AUG-H1- Ψ (three-way-1) junction (Fig. 3, B and D). A227 to U291 forms an extended DIS hairpin with two internally stacked but nonpaired guanines (G272 and G273) and a G240(syn):G278(anti)-G241(anti) base triple. Sequentially stacked pyrimidines (U230*U288 and C231*C287) exhibit broad line widths indicative of millisecond time scale conformational exchange (Fig. 3E). These residues appear to function as a flexible hinge that connects the extended DIS hairpin with the tandem three-way junction (Fig. 3D). U307 to G330 forms an extended ψ -hairpin structure that contains three noncanonical base pairs [G310(anti)*A327(anti), G328*U309, G329*U308], a stacked A-A bulge

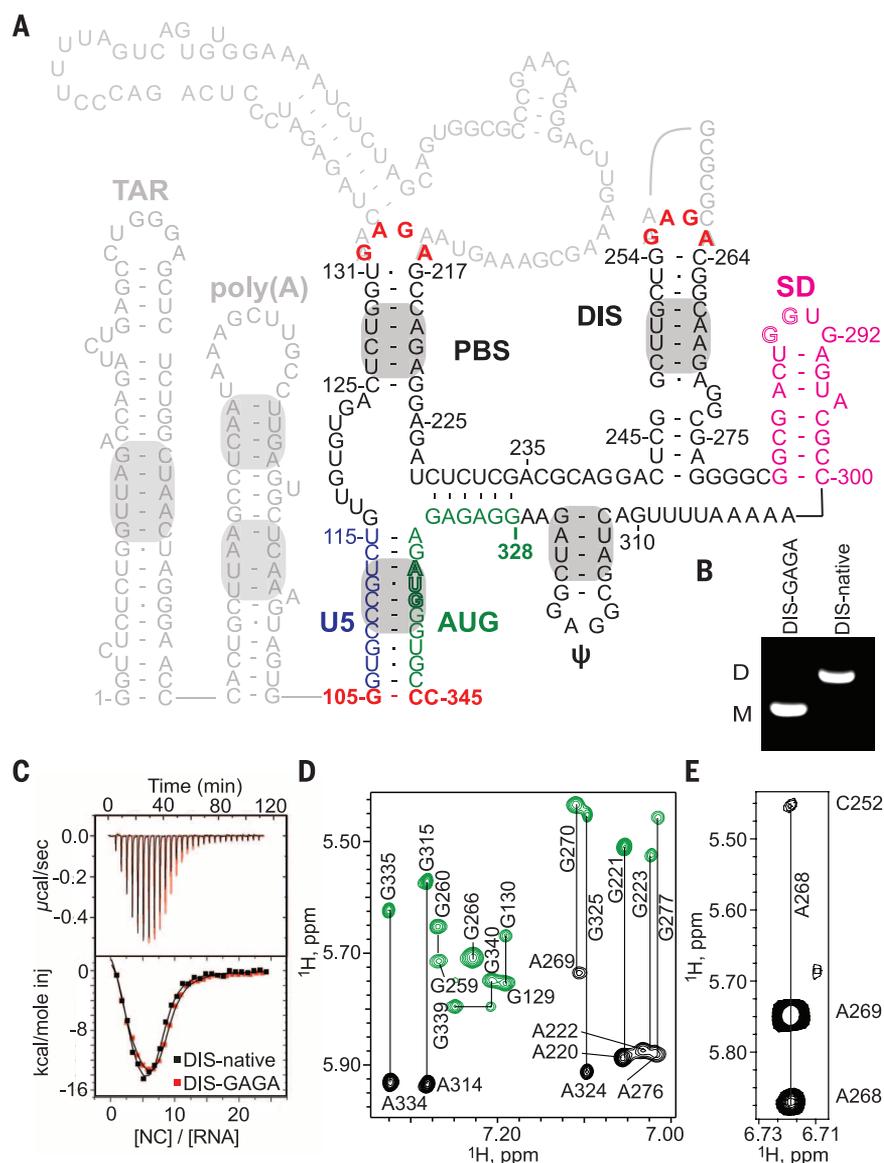


Fig. 1. HIV-1 5' leader and Ψ^{CES} RNA construct. (A) Predicted secondary structure of the HIV-1 (NL4-3 strain) 5' leader (16); gray shading denotes elements detected in the intact leader by NMR (13, 15); dark letters denote Ψ^{CES} (nonnative residues colored red; see text). (B and C) Substitution of the native DIS loop residues (DIS-native) by GAGA (DIS-GAGA) prevents dimerization (B) but does not affect NC binding (C). ppm, parts per million. (D) Representative NOESY spectra for G⁸A- Ψ^{CES} (black) and G- Ψ^{CES} (green); lines connect H₈ (vertical labels) and H₁' (horizontal labels) signals. (E) Representative very-long-range NOE (A268-H₂ to C252-H₁'; ~ 7 Å separation) obtained for A^{2r}C^r- Ψ^{CES} .

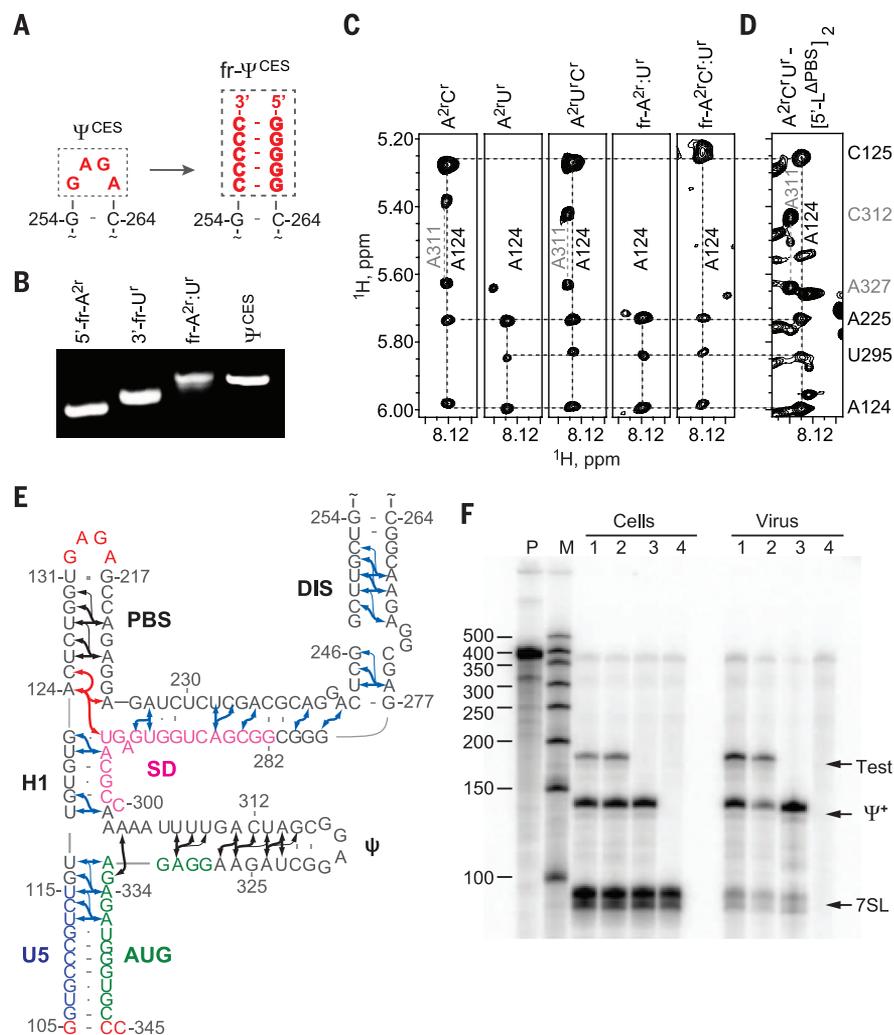


Fig. 2. Fragmentation-based ^2H -edited NMR approach and observed Ψ^{CES} secondary structure. (A) The DIS loop of Ψ^{CES} served as the fragmentation site and was substituted by a stretch of intermolecular G-C base pairs. (B) Fragment-annealing efficiency as measured by native agarose gel electrophoresis. (C) The 2D NOESY spectra of uniformly labeled $\text{A}^{2r}\text{C}'\text{r}$ -, $\text{A}^{2r}\text{U}'\text{r}$ -, and $\text{A}^{2r}\text{C}'\text{r}:\text{U}'\text{r}$ - Ψ^{CES} and segmentally labeled $\text{fr-A}^{2r}:\text{U}'\text{r}$ - and $\text{fr-A}^{2r}\text{C}'\text{r}:\text{U}'\text{r}$ - Ψ^{CES} samples used to make long-range NOE assignments. (D) Similarities in NOESY spectra obtained for $\text{A}^{2r}\text{C}'\text{r}:\text{U}'\text{r}$ -labeled $[\text{5}'\text{-L}^{\text{APBS}}]_2$ and Ψ^{CES} confirm that the tandem three-way junction structure is present in both constructs. (E) NMR-derived secondary structure of Ψ^{CES} . Black and blue arrows denote A-H₂ NOEs observable in Ψ^{CES} and $\text{fr-}\Psi^{\text{CES}}$ samples, respectively; red arrows highlight NOEs shown in (C) and (D); thin arrows denote very-long-range NOEs. (F) Packaging of native HIV-1_{NL4-3} 5'-L and 5'-L^{APBS} RNAs under competition conditions assayed by means of ribonuclease protection. P, undigested probe; M, RNA sizes marker. Lanes 1 and 2 show native HIV-1_{NL4-3} helper versus test vectors containing 5'-L^{APBS} (1) or native HIV-1_{NL4-3} (2). Lane 3 contains HIV-1_{NL4-3} helper expressed without test RNA. Lane 4 is mock transfected cells. Samples obtained from transfected cells (Cells) or viral-containing media (Virus) are indicated. Bands corresponding to host 7SL RNA, HIV-1_{NL4-3} helper RNA (Ψ^+), and copackaged test RNAs (Test) are labeled.

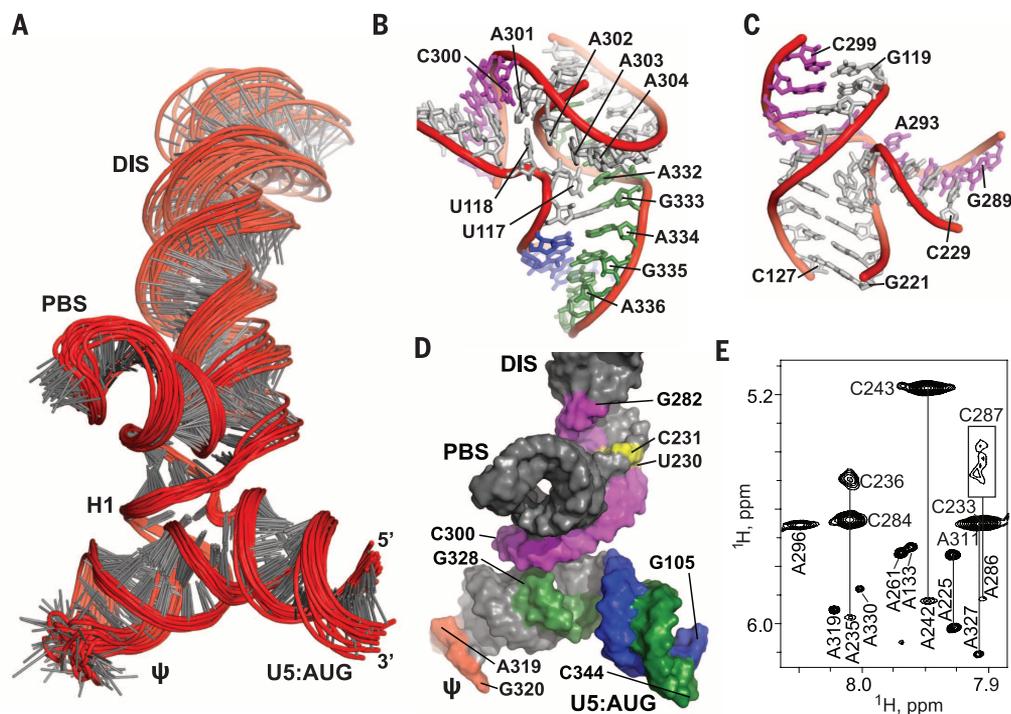


Fig. 3. Three-dimensional structure of Ψ^{CES} . (A) Ensemble of 20 refined structures (residues 105 to 344 shown). (B and C) Expanded views of the (B) three-way-1 and (C) three-way-2 junctions. (D) Surface representation of Ψ^{CES} highlighting U5 (blue); AUG (green) base pairing and the integral participation of SD residues (pink) in the tandem three-way junction structure. (E) Severe line broadening indicative of slow (millisecond) conformational averaging was observed for stacked, mismatched pyrimidines in the extended DIS stem [yellow in (D); broadened C287-H₁' signal boxed in (E)]. NOE patterns and sharp NMR signals also indicate that the ψ hairpin loop is unstructured [red in (D)].

[A311(anti)-A326(anti)] (Fig. 2E), and a flexible GAGG loop (Fig. 3D). Adenosines A302 to A305 exhibit pseudo A-form stacking but are not base-paired (Fig. 3B), which supports proposals that genomic adenosine enrichment occurs primarily at non-base-paired sites (23). A302 and A303 also make A-minor contacts with the U5:AUG helix (Fig. 3B).

To determine whether the tandem three-way junction is evolutionarily conserved, we analyzed published HIV-1 leader sequences that contained full coverage of the 5' untranslated region (278 total sequences). Representatives from B, C, and F1 subtypes were included in the analysis (18). Of the 48 base-paired nucleotides at or near the three-way junction, 31 were either strictly (16 sites) or very highly (>99%, 15 sites) conserved, and 13 displayed high (90.2% to 98.9%) identity (table S2). Only 11 of 126 substitutions resulted in loss of base pairing. The remaining four sites—A²²⁷, G²⁷⁹, A²⁸⁶, and U²⁸⁸—exhibited significant variation, ranging from 12% (U²⁸⁸) to 50.3% (A²²⁷). Most changes mapped to terminal branches of the Ψ^{CES} phylogeny. Thus, the tandem three-way junction structure is highly conserved, and the rare variations that disrupt base pairing are due to transient polymorphisms.

The PBS, DIS, and ψ helices of Ψ^{CES} are consistent with models derived from nucleotide reactivity experiments (4), but the SD structure differs significantly. Recent in-gel chemical probing of resolved monomeric and dimeric leader RNAs (24), and probing studies under solution conditions that favor either the monomeric or dimeric species (25), showed that SD loop residues are relatively unreactive in the dimeric RNA, consistent with the Ψ^{CES} structure. Pseudo-free energy calculations indicate that the in-gel reactivity data for the dimeric leader (24) are in better agreement with the Ψ^{CES} NMR structure than the proposed model [~25% lower experimental pseudo-free energy (18); fig. S4]. These findings support proposals that variations in structure predictions are at least partly due to site-specific structural heterogeneity associated with the monomer-dimer equilibrium (13, 24).

HIV-1 NC binds with high affinity to oligonucleotides that contain exposed guanines (4, 26, 27). Ψ^{CES} contains five unpaired Gs (excluding the nonnative GAGA tetraloops), a GGG base triple in the DIS stem, and five additional guanine mismatches clustered at or near the two three-way junctions (G*U, G*A, or G*G) that could serve as NC binding sites (Fig. 4A). Potential contributions of these "junction guanines" to NC binding were tested by isothermal titration calorimetric studies of G-to-A-substituted Ψ^{CES} RNAs. Free energy calculations indicate that these substitutions, which include three G*U to A-U substitutions, should not alter the secondary structure of the RNA (18). Replacement of the ψ GGAG loop by GAAA eliminated one NC binding site, as expected (27), and substitution of the three-way-1 junction guanines by adenosines (G116A/G333A/G328A/G329A/G331A) eliminated three additional NC sites (Fig. 4B). Mutation of the unpaired (G226, G292, and G294) and mismatched (G224)

three-way-2 junction guanines to adenosines eliminated one NC binding site (Fig. 4B). The influence of these guanines on RNA encapsidation was evaluated by using a competitive in situ RNA packaging assay. Human embryonic kidney 293T cells were co-transfected with plasmids that produce vector RNAs containing the wild-type (Ψ^+ , which also encodes for viral proteins) and mutant (Test) leader sequences (18). When coexpressed at similar levels, Ψ^+ and Test vector RNAs with native leader sequences were packaged into HIV-1 virus-like particles with similar efficiencies (Fig. 4C). In contrast, significant packaging defects were observed upon G-to-A mutation of the three-way-2 junction guanines (17% \pm 2%), the ψ -loop and three-way-1 junction guanines (10% \pm 2%), and all junction and ψ -loop guanines (5% \pm 1%) (Fig. 4C). Our findings indicate

that the tandem three-way junction serves as a scaffold for exposing clusters of unpaired or weakly paired junction guanines, thereby enabling their binding to the zinc knuckle domains of NC.

The Ψ^{CES} structure explains biochemical, nucleotide reactivity, and phylogenetic results and suggests a mechanism by which the 5' leader structure regulates translation and splicing (4). In vitro translational activity and chemical reactivity of the AUG residues are suppressed upon dimerization (28), and this can be attributed to sequestration of the 5' end of the *gag* open reading frame within the three-way-1 junction (Fig. 3D). Enhanced in vitro translational activity caused by mutations immediately downstream of the major splice donor site (Δ A296/A301U and A293C/U295C/ Δ G298) can be explained by

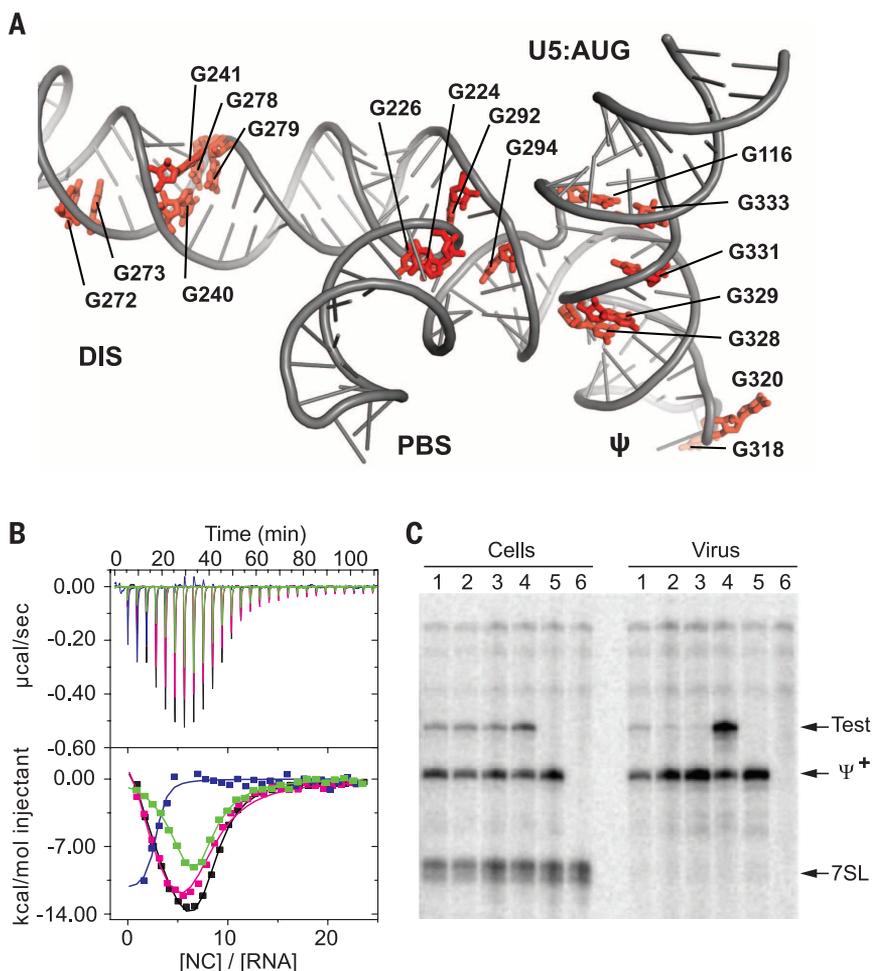


Fig. 4. Junction guanines mediate NC binding and packaging. (A) Ψ^{CES} contains 17 unpaired or weakly paired guanines (red) that serve as potential NC binding sites. (B) Mutation of the three-way-2 (green) or ψ (magenta) guanines to adenosines modestly reduces NC binding ($N = 7.0 \pm 0.3$ and 7.0 ± 0.5 , respectively) relative to wild-type Ψ^{CES} (black; $N = 8.0 \pm 0.3$). Mutation of ψ and three-way-1 guanines to adenosines (blue) severely inhibits high-affinity NC binding ($N = 2.0 \pm 0.1$). (C) Competitive packaging of HIV-1_{NL4-3} vectors containing native and mutant 5' leader sequences, assayed by means of ribonuclease protection. Lanes 1 to 4 are native HIV-1_{NL4-3} versus test vectors containing 5'-L^{3way2-G/A} (1), 5'-L^{3way1-G/A} (2), 5'-L^{3way1,2-G/A} (3), and 5'-L (4). Lane 5 is HIV-1_{NL4-3} helper expressed without test RNA. Lane 6 is mock transfected cells. Samples obtained from transfected cells (Cells) or viral containing media (Virus) are indicated. Bands corresponding to host 7SL RNA, HIV-1_{NL4-3} helper RNA (Ψ^+), and copackaged test RNAs (Test) are labeled.

destabilization of the H1 helix and, for $\Delta A296/A301U$, stabilization of the SD hairpin (29), both of which should favor the monomer. Mutations in AUG that inhibit genome dimerization and suppress packaging (30, 31) are expected to disrupt base pairing in the U5:AUG helix and ψ -hairpin stem, thereby destabilizing the tandem three-way junction structure required for Gag binding. In vitro splicing activity is also attenuated by dimerization (12, 32), and this can be attributed to sequestration of the major splice-site recognition sequence within the three-way-2 junction. Antisense oligonucleotides with complementarity to the SD loop inhibit dimerization, and this is likely due to their ability to competitively block formation of the tandem three-way junction (25).

The Ψ^{CES} structure also explains the exquisite selectivity of HIV-1 to package its unspliced genome (1, 2). Residues immediately downstream of the major splice site are base-paired within the H1 helix and are thus integral to the formation of the tandem three-way junction structure. Although unspliced and spliced HIV-1 mRNAs contain identical 5' sequences (G1 to G289), differences in spliced mRNA sequences derived from 3' exons would preclude formation of the packaging competent junction structure. Similarly, because SD appears to exist as a hairpin in the monomeric, unspliced 5' leader (12), it is likely that monomeric genomes are also ignored during virus assembly because they do not adopt the tandem three-way junction structure.

Compared with the proteins of HIV-1, structural information for the viral nucleic acids is sparse. RNAs in general are vastly underrepresented in the structural data banks (99,000 proteins versus 2700 RNA structures), partly because of NMR technical challenges and difficulties obtaining suitable crystals for x-ray diffraction (19, 20). The fr-RNA 2H -edited NMR approach enables efficient segmental labeling without requiring enzymatic ligation. Given the ubiquity of hairpin elements that can serve as fragmentation or annealing sites, this method should be generally applicable to modest-sized RNAs (~160 nucleotides).

REFERENCES AND NOTES

- J. M. Coffin, S. H. Hughes, H. E. Varmus, Eds., *Retroviruses* (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1997).
- R. Berkowitz, J. Fisher, S. P. Goff, *Curr. Top. Microbiol. Immunol.* **214**, 177–218 (1996).
- J. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13535–13540 (2009).
- K. Lu, X. Heng, M. F. Summers, *J. Mol. Biol.* **410**, 609–633 (2011).
- N. Jouvenet, S. M. Simon, P. D. Bieniasz, *J. Mol. Biol.* **410**, 501–511 (2011).
- M. Kuzembayeva, K. Dille, L. Sardo, W.-S. Hu, *Virology* **454–455**, 362–370 (2014).
- S. Schwartz, B. K. Felber, D. M. Benko, E. M. Fenyó, G. N. Pavlakis, *J. Virol.* **64**, 2519–2529 (1990).
- O. A. Nikolaitchik *et al.*, *PLOS Pathog.* **9**, e1003249 (2013).
- A. M. Lever, *Adv. Pharmacol.* **55**, 1–32 (2007).
- J.-C. Paillart, M. Shehu-Xhilaga, R. Marquet, J. Mak, *Nat. Rev. Microbiol.* **2**, 461–472 (2004).
- J. Greatorex, *Retrovirology* **1**, 22 (2004).
- T. E. M. Abbink, B. Berkhout, *J. Virol.* **82**, 3090–3098 (2008).
- K. Lu *et al.*, *Science* **334**, 242–245 (2011).
- T. E. M. Abbink, B. Berkhout, *J. Biol. Chem.* **278**, 11601–11611 (2003).
- X. Heng *et al.*, *J. Mol. Biol.* **417**, 224–239 (2012).
- K. A. Wilkinson *et al.*, *PLOS Biol.* **6**, e96 (2008).
- S. B. Kutluay *et al.*, *Cell* **159**, 1096–1109 (2014).
- Information on materials and methods is available on Science Online.
- K. Lu, Y. Miyazaki, M. F. Summers, *J. Biomol. NMR* **46**, 113–125 (2010).
- O. Duss, P. J. Lukavsky, F. H. Allain, *Adv. Exp. Med. Biol.* **992**, 121–144 (2012).
- L. M. Koharudin, A. M. Bonvin, R. Kaptein, R. Boelens, *J. Magn. Reson.* **163**, 228–235 (2003).
- S. Barton, X. Heng, B. A. Johnson, M. F. Summers, *J. Biomol. NMR* **55**, 33–46 (2013).
- F. J. van Hemert, A. C. van der Kuyl, B. Berkhout, *RNA Biol.* **10**, 211–215 (2013).
- J. C. Kenyon, L. J. Prestwood, S. F. Le Grice, A. M. Lever, *Nucleic Acids Res.* **41**, e174 (2013).
- J. Deforges, N. Chamond, B. Sargueil, *Biochimie* **94**, 1481–1489 (2012).
- T. L. South, M. F. Summers, *Protein Sci.* **2**, 3–19 (1993).
- R. N. De Guzman *et al.*, *Science* **279**, 384–388 (1998).
- F. Baudin *et al.*, *J. Mol. Biol.* **229**, 382–397 (1993).
- T. E. Abbink, M. Ooms, P. C. Haasnoot, B. Berkhout, *Biochemistry* **44**, 9058–9066 (2005).
- D. T. Poon, E. N. Chertova, D. E. Ott, *Virology* **293**, 368–378 (2002).
- O. Nikolaitchik, T. D. Rhodes, D. Ott, W.-S. Hu, *J. Virol.* **80**, 4691–4697 (2006).
- J. A. Jablonski, E. Buratti, C. Stuaní, M. Caputi, *J. Virol.* **82**, 8038–8050 (2008).

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of General Medical Sciences (NIGMS, R01 GM42561 to M.F.S. and A.T. and P50 GM 103297 to M.S., B.J., and D.A.C.). S.B., N.C.B., and S.M. were supported by a NIGMS grant for enhancing minority access to research careers (MARC U*STAR 2T34 GM008663), and S.B., J.S., N.C.B., and S.M. were supported by an HHMI undergraduate education grant. We thank the HHMI staff at UMBC for technical assistance and B. Rife (University of Florida) for advice regarding the phylogenetic analysis. The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, National Institute of Allergy and Infectious Diseases, NIH: pNL4-3 from M. Martin. Atomic coordinates have been deposited into the Protein Data Bank with accession code 2NIQ. NMR chemical shifts and restraints have been deposited into the Biological Magnetic Resonance Bank with accession code 25571.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6237/917/suppl/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References (33–52)

14 February 2015; accepted 22 April 2015
10.1126/science.aaa9266

EVOLUTION

Systematic humanization of yeast genes reveals conserved functions and genetic modularity

Aashiq H. Kachroo,¹ Jon M. Laurent,¹ Christopher M. Yellman,¹ Austin G. Meyer,^{1,2} Claus O. Wilke,^{1,2,3} Edward M. Marcotte^{1,2,4*}

To determine whether genes retain ancestral functions over a billion years of evolution and to identify principles of deep evolutionary divergence, we replaced 414 essential yeast genes with their human orthologs, assaying for complementation of lethal growth defects upon loss of the yeast genes. Nearly half (47%) of the yeast genes could be successfully humanized. Sequence similarity and expression only partly predicted replaceability. Instead, replaceability depended strongly on gene modules: Genes in the same process tended to be similarly replaceable (e.g., sterol biosynthesis) or not (e.g., DNA replication initiation). Simulations confirmed that selection for specific function can maintain replaceability despite extensive sequence divergence. Critical ancestral functions of many essential genes are thus retained in a pathway-specific manner, resilient to drift in sequences, splicing, and protein interfaces.

The ortholog-function conjecture posits that orthologous genes in diverged species perform similar or identical functions (1). The conjecture is supported by comparative analysis of gene-expression patterns, genetic interaction maps, and chemogenomic profiling

(2–6), and it is widely used to predict gene function across species. However, even if two genes perform similar functions in different organisms, it may not be possible to replace one for the other, particularly if the organisms are widely diverged. The extent to which deeply divergent orthologs can stand in for each other, and which principles govern such functional equivalence across species, is largely unknown.

In this study, we systematically addressed these questions by replacing a large number of yeast genes with their human orthologs. Humans and the baker's yeast *Saccharomyces cerevisiae* diverged from a common ancestor approximately 1 billion years ago (7). They share several thousand

¹Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA. ²Center for Computational Biology and Bioinformatics, University of Texas at Austin, Austin, TX 78712, USA. ³Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA. ⁴Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA.

*Corresponding author. E-mail: marcotte@cmb.utexas.edu

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 21, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6237/917.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/05/20/348.6237.917.DC1.html>

This article **cites 47 articles**, 18 of which can be accessed free:

<http://www.sciencemag.org/content/348/6237/917.full.html#ref-list-1>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>